Accident Analysis and Prevention 83 (2015) 1–17

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

A framework of boundary collision data aggregation into neighbourhoods

Ge Cui[a], Xin Wang[a,*], Dae-Won Kwon[b]

[a] Department of Geomatics Engineering, University of Calgary, Calgary, Canada
[b] Office of Traffic Safety, City of Edmonton, Edmonton, Canada

ARTICLE INFO

Article history:
Received 20 January 2015
Received in revised form 4 May 2015
Accepted 5 June 2015
Available online xxx

Keywords:
Boundary collision aggregation
Entropy
Boundary zone
Collision density ratio

ABSTRACT

A large portion of the total number of motor collisions can be boundary collisions; therefore, exaggerated or underestimated numbers for boundary collisions aggregated into neighbourhoods may hamper road safety analyses and management. In this paper, we propose a systematic framework for boundary collision aggregation. First, an entropy-based histogram thresholding method is utilized to determine the boundary zone size and identify boundary collisions. Next, the collision density probability distribution is then established, based on the collisions in each neighbourhood. Last, an effective boundary collision aggregation method, called the collision density ratio (CDR), is used to aggregate boundary collisions into neighbourhoods. The proposed framework is applied to collision data in the City of Edmonton for a case study. The experimental results show that the proposed entropy-based histogram thresholding method can identify boundary collision with the high precision and recall, and the proposed CDR method is more effective than the existing methods, the half-to-half ratio method and the one-to-one ratio method, to aggregate boundary collisions into neighbourhoods.

1. Introduction

Road traffic accidents are social and public health challenges, as they almost always result in injuries and/or fatalities (World Health Organization, 2013). The World Health Organization reports that road collisions, as the ninth leading cause of death in 2004, will be ranked as the fifth leading cause in 2030 (World Health Organization, 2013). It has been estimated that over one million people are killed each year in road collisions, which is equivalent to 2.1% of the annual global mortality, resulting in an estimated social cost of $518 billion (Peden et al., 2004).

Previous traffic safety studies have shown that the occurrences of motor vehicle accidents are rarely random in space and time; therefore, the macro-level analysis of collision data is a substantial component of traffic planning and traffic management. Examples include LaScala et al. (2001) conducting a geostatistical analysis to examine the relationship of neighbourhood characteristics to alcohol-related pedestrian injury collisions, the use of macro-level collision prediction models (CPMs) in road safety evaluation and planning (Lovegrove and Sayed, 2006), and some studies showing that neighbourhood street patterns have a significant impact on traffic collision frequency (Rifaat et al., 2009). A critical step in conducting a macro-level analysis of collision data in road safety is the effective aggregation of collisions into neighbourhoods. The aggregation of collision data has a large impact on traffic analysis and management.

Boundary collisions are motor accidents that occur on the boundaries of neighbourhoods. Boundaries are usually defined by conspicuous natural or artificial ground objects, such as main roads, rail lines, trails and rivers. As boundaries of neighbourhoods are often main roads where most collisions happen, boundary collisions can account for a large proportion of the total collisions (Siddiqui and Abdel-Aty, 2012; Wang et al., 2012; Lee et al., 2014). Therefore, traffic analysis and management are considerably affected by these boundary collisions: this phenomenon is referred to as the *boundary effect*.

Boundary collisions can, however, be difficult to identify after the digitalization and geocoding process, for the following reasons:

- Boundaries may not be coincident with the corresponding roads on the map, which leads to boundary collisions not located on the boundary lines, and
- In the representation of roadways in the features of a geographic information system (GIS) where, for example, roadways with a width of 10 m are often represented as single lines, collisions may deviate from the roads.

* Corresponding author.
E-mail address: xcwang@ucalgary.ca (X. Wang).

http://dx.doi.org/10.1016/j.aap.2015.06.003
0001-4575/

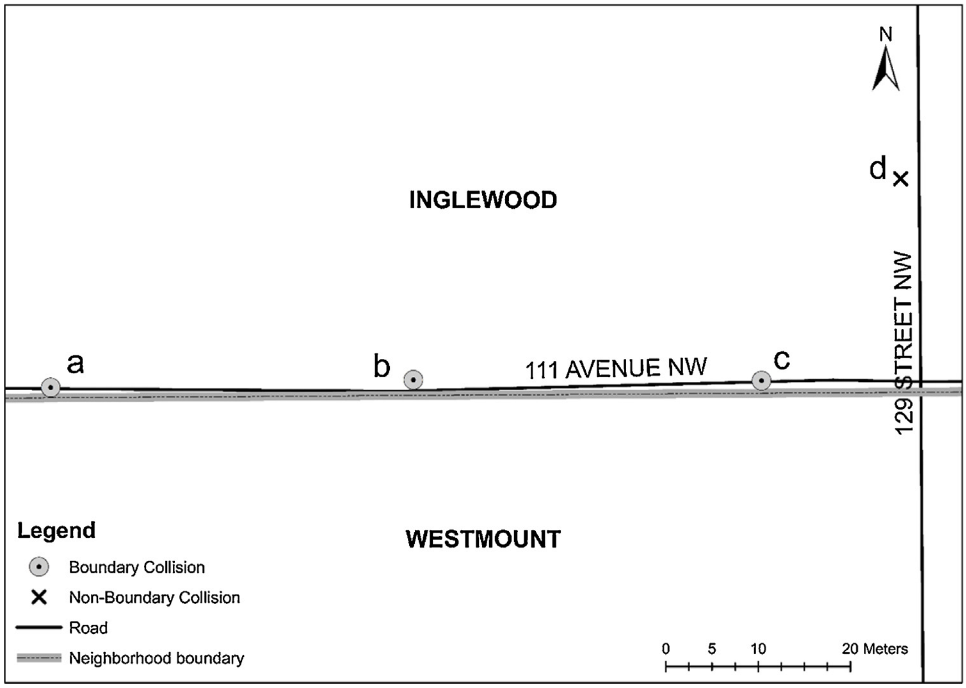

Fig. 1. Boundary collision identification.

Fig. 1 illustrates these two cases by using real collision data from the City of Edmonton in Alberta, Canada. In the figure, 111 Avenue NW is a roadway served as the boundary between the neighbourhoods "INGLEWOOD" and "WESTMOUNT". Three boundary collisions 'a', 'b' and 'c' are on the boundary road 111 Avenue NW and one non-boundary collision 'd' is close to 129 Street NW. In Fig. 1, the road and the boundary do not match after digitalization. Moreover, after the digitalization, though collision 'b' is located very close to the boundary, it is neither on the road nor on the boundary.

The manual inspection of boundary collisions is time-consuming and requires significant human resources. A better solution is the generation of a boundary zone to include nearby collisions. The boundary zone is a buffer centred at the boundaries of neighbourhoods. Collisions located within a boundary zone are assumed as boundary collisions; and, boundary collisions can then be assigned to the neighbourhoods through aggregation methods. Fig. 2 shows a 6-m boundary zone for the example in Fig. 1. In Fig. 2, the boundary zone contains all three collisions 'a', 'b' and 'c' as boundary collisions. The collision 'd' is beyond the boundary zone and treated as a non-boundary collision.

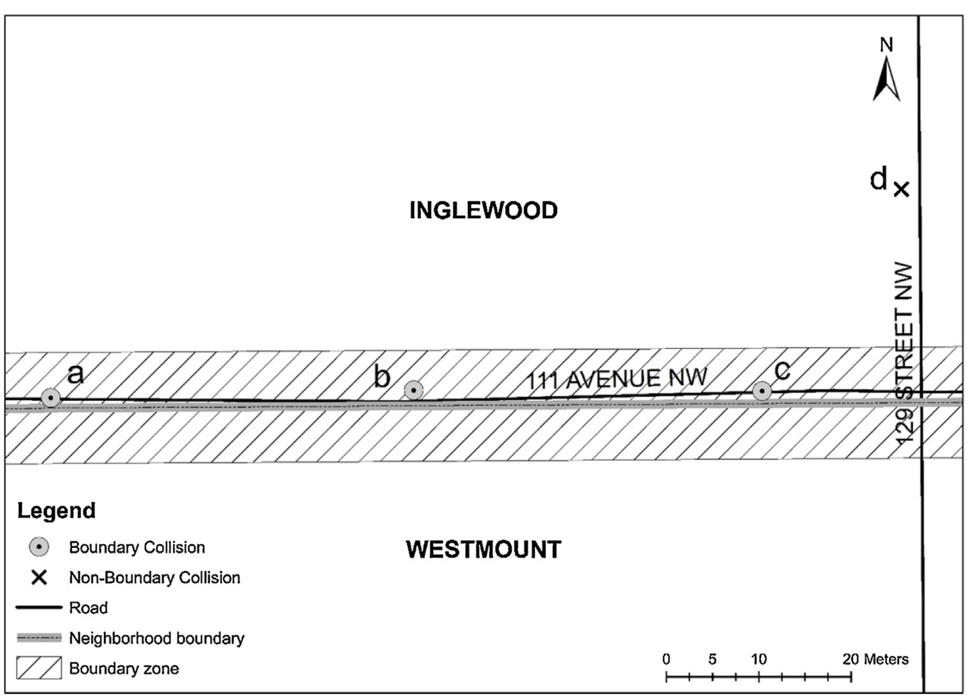

Fig. 2. Boundary zone.